

基于符号化表示相似度度量的卫星多元 工程参数异常检测*

宋华婷^{1,2} 刘玉荣¹

1(中国科学院国家空间科学中心 北京 100190)

2(中国科学院大学 北京 100049)

摘 要 随着卫星系统复杂程度的日益增加,综合分析卫星多元参数之间的相关性异常对于卫星安全运行和空间任务的正确执行具有重要意义。利用某卫星工程参数数据,基于符号聚合近似算法(SAX),研究卫星多元工程参数的异常检测问题,解决了当前异常检测方法中多元参数融合时不考虑上下文信息造成信息丢失的问题,实现多元参数有效融合,形成一种优化的基于快速动态时间规整算法(Fast-DTW)的异常检测算法。研究结果表明,模型在某卫星电源子系统的异常检测过程中,recall, precision 和 F1 score 分别为 0.947, 0.9 和 0.923,能够实际应用于卫星异常检测,提高卫星在轨运行的安全性。

关键词 异常检测, 卫星工程参数, 符号聚合近似, 快速动态时间规整 (Fast-DTW)

中图分类号 TP391

Multivariate Engineering Parameter Anomaly Detection of the Satellite Based on Similarity Metric of the Symbolic Representation

SONG Huating^{1,2} LIU Yurong¹

1(National Space Science Center, Chinese Academy of Sciences, Beijing 100190)

2(University of Chinese Academy of Sciences, Beijing 100049)

Abstract The complexity of the satellite system is increasing, and the comprehensive analysis of the abnormal correlation between the satellite with multiple parameters is important for the safe operation of the satellite and the correct execution of space tasks. Aiming at the characteristics of large amount of engineering parameter data, high parameter correlation and pseudo-period, a satellite multivariate engi-

* 中国科学院战略性先导科技专项(A类)空间科学(二期)地面支撑系统科学卫星任务运控技术(XDA15040100)和空间先导专项科学卫星在轨运行维护科学卫星运控管理与服务(E02215A01S)共同资助

2022-01-12 收到原稿, 2022-11-10 收到修定稿

E-mail: songhuating19@mailsucas.ac.cn

neering parameter anomaly detection method based on symbolized representation similarity measurement is proposed. Using real satellite engineering parameter data, based on Symbol Aggregation Algorithm (SAX), this paper studies the anomaly detection problem of satellite multiple engineering parameters, solves the problem of the context information, realizes effective fusion, and forms an optimized anomaly detection algorithm based on similarity measurement Fast-DTW algorithm. Results show that during the abnormality detection process, the recall, precision and F1 score are 0.947, 0.9 and 0.923 respectively in the real satellite power subsystem, and the algorithm can be actually used in satellite anomaly detection to improve the safety of satellite in-orbit operation.

Key words Anomaly detection, Satellite engineering parameters, Symbolic aggregate approximation, Fast Dynamic Time Warping (Fast-DTW)

0 引言

卫星是集成了电、光、热、力、磁等多种技术手段的复杂综合系统。随着航天技术和航天应用的发展,卫星系统复杂程度不断增加。在轨运行过程中,由于元器件老化、恶劣的空间环境等原因,卫星会发生在轨异常,影响其在轨运行任务。根据对中国 7 颗在轨低轨卫星的统计,在 2008 至 2012 年发生了 183 例在轨异常事件^[1],严重影响卫星运行的可靠性和空间任务执行的安全性。卫星运行过程中产生的参数数据,是地面测运控中心判断卫星运行状态和工作性能的重要依据。基于工程参数的卫星异常检测,对于提高卫星在轨运行的可靠性、安全性和有效性具有重要作用,已成为航天领域重点研究方向之一。

卫星工程参数异常可分为点异常、上下文异常和组合异常。点异常可认为是在数据集中偏离大部分数据的异常数据,上下文异常可认为是不同工作模式下参数呈现不同的状态检测,组合异常是一组具有密切相关性的多元工程数据发生关联变化呈现的异常。

卫星工程参数具有高维、时序、数据量大和伪周期的特性^[2]。目前针对卫星异常检测的研究主要分为两类^[3]: 基于人工结合阈值判读、规则、物理模型的方法和基于数据驱动的方法。

Hundman 等^[4]采用手动设置阈值的方式进行实际航天器长管系统检测; Zhang 等^[5]将故障树分析法与专家系统相结合建立了故障诊断框架,用于实际任务中有效载荷状态故障诊断; Figueroa 等^[6]根据航天器的系统组成、工作原理等建立物理模型,将模型预测结果与实际工程数据进行比较,实现异常检测。以

上研究方法具有可解释性较强、误检率低、实施简单的特点,针对单参数的超限异常以及复杂性较低的少量多参数相关异常能够有效进行检测。但是,建立专家知识库以及物理模型过程中成本较高,对于差别较大的航天器系统,需要重新构建知识库和模型。

基于数据驱动的时间序列异常检测方法,通过对工程数据建模以识别异常模式。可大致分为基于相似性度量、基于机器学习以及基于状态估计的方法。Berndt 等^[7]提出了一种基于动态时间规整(DTW)的异常检测方法,用于计算长度不一致的时间序列相似性度量; Tian^[8]提出了基于深度学习及 GPU 计算对航天器故障进行判定的技术,实现航天器故障自动学习与检测,对未知的故障表现良好; Gu^[9]建立 ARI-MA-SVR 组合模型对真实遥测数据进行趋势预测,提高了预测精度。基于数据驱动的方法不依赖于专家经验,在专家知识不足的情况下,仍能表现出较好的异常检测能力。

复杂卫星系统参数之间关系复杂,只对单元进行异常检测难以满足卫星实际状态监测的需求。Wang 等^[10]将符号聚合近似(SAX)和 KNN 算法用于机场噪声异常检测, SAX 大大降低维度, KNN 减小了 SAX 带来的信息损失; Ruan 等^[11]提出一种趋势的 SAX 算法,通过描述序列均值和斜率构建分类器,用于卫星时序数据分类。针对传统 DTW 算法只关注单点信息而忽略了上下文信息的问题, Wang 等^[12]提出了一种结合形状特征及其上下文的多维 DTW 算法。Du 等^[13]针对伪周期时序序列,基于滑动窗口和极值提出了一种基于均序列动态生成模型的子序列异常检测方法。

现有的卫星多元工程参数异常检测方法归纳如下。

(1) 多元特征提取采用的主成分分析法(PCA)属于压缩式降维方法,不适用于提取高相关性参数特征。

(2) 特征融合对单元参数主要基于经典的相似性或距离度量方法进行计算,往往忽视了多元参数能否有效融合。

(3) 常用的子序列划分算法不适用于具有伪周期性的卫星时序数据。

综上,针对卫星工程参数伪周期、数据量大和参数相关性高的特性,这里设计了一个分段方法用于工程参数序列预处理。在 SAX 和 N-Gram 算法的基础上,使用滑动窗口结合上下文信息对符号化过程和特征提取过程进行改进,并与快速动态时间规整算法 Fast-DTW 相结合,建立了面向相似性的异常检测模型,形成一种基于符号化表示相似性度量的卫星多元工程参数异常检测方法。

1 定义

1.1 时间序列

m 元时间序列数据集 A 可以表示为

$$A = (T_1, T_2, \dots, T_i, \dots, T_m).$$

其中, T_i 表示第 i 元时间序列数据。长度为 n 的第 i 元时间序列数据可以表示为

$$T_i = (t_{i1}, t_{i2}, \dots, t_{ij}, \dots, t_{in}).$$

其中, t_{ij} 表示数据集中第 i 元参数在时刻点 j 的数值。

1.2 子序列

使用窗口大小为 l 的滑动窗口进行子序列提取,每次移动步长为 p 。其可以表示为

$$S_{\text{eg}}(T, l, p) = \{S_1, S_2, \dots, S_{(n-l)/p+1}\}.$$

其中,每个子序列表示为

$$S_i = \begin{cases} (t_{1+(i-1)\times p}, \dots, t_{(i-1)\times p+l}), & 1 \leq i \leq (n-l)/p; \\ (t_{1+(i-1)\times p}, \dots, t_n), & i = (n-l)/p + 1. \end{cases}$$

1.3 截断

将长度为 l_{en} 的序列 S_i 划分为 k 个截断 $\bar{s}_j$ ($j = 1, 2, \dots, k$),得到 $\bar{S}_i = (\bar{s}_1, \bar{s}_2, \dots, \bar{s}_k)$,取截断内数据的均值表示每个截断,有

$$\bar{s}_j = \frac{k}{l} \sum_{w=\frac{l_{\text{en}}}{k}(j-1)+(i-1)\times p+1}^{\frac{l_{\text{en}}}{k}j} t'_w.$$

1.4 符号化

选择字母表大小为 φ ($\varphi > 2$),即表示序列的字母数,使用 $\varphi - 1$ 个断点将高斯分布空间划分为 φ 个等概率区域,可表示为 $\{(-\infty, \beta_1), (\beta_1, \beta_2), \dots, (\beta_{\varphi-1}, +\infty)\}$ 。表1列出了字母表数为3~7的断点查找表^[14]。

根据字母表大小在断点查找表中查找区间的系列分裂点 β_i ,将均值表示为对应的字母。 $\varphi = 4$ 和 $\varphi = 5$ 时均值和对应的字母映射关系如图1所示。

2 算法

2.1 基于极值的伪周期时间序列分段算法

2.1.1 卫星时序数据分析

基于相似度度量的异常检测方法,需要将参数数据划分为呈现相似规律的子序列。划分过程是通过寻找某种规律将参数序列分段。

对于具有周期特征的卫星工程参数,由于卫星受到太阳、轨道、SAA 等多方面的影响,往往呈现典型

表1 字母数为3~7的断点查找表

Table 1 Breakpoint lookup table with alphabetically numbered 3~7

	3	4	5	6	7
β_1	-0.43	-0.67	-0.84	-0.97	-1.07
β_2	0.43	0	-0.25	-0.43	-0.57
β_3	—	0.67	0.25	0	-0.18
β_4	—	—	0.84	0.43	0.18
β_5	—	—	—	0.97	0.57
β_6	—	—	—	—	1.07

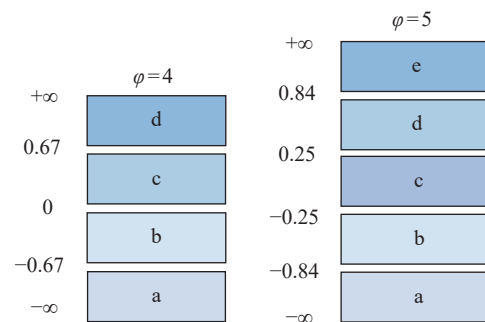


图1 $\varphi = 4$ 和 $\varphi = 5$ 时字母映射关系

Fig. 1 Latter mapping relationship when $\varphi = 4$ and $\varphi = 5$

的伪周期性。其特点是参数数据存在大量的重复趋势,整体相似但局部有差别且周期长度不相同。

图2是某卫星电源分系统充电电流时序数据,可以看出,卫星时序数据具有明显的周期性,并且整体变化趋势相似,局部有细小差别,呈现典型的伪周期性。每个子序列的长度不相同,但是内部存在典型的跳变值,体现子序列的总体特征。传统的固定窗口子序列分段算法不能很好地解决卫星参数伪周期特性问题。图3给出的是图2数据序列基于传统的固定窗口子序列分段算法得到的分段结果,可以看出,依据简单的固定窗口进行分段时,受到序列伪周期特性的影响,容易出现不合理的分段结果。

由于卫星参数数据周期规律呈现一种以卫星运行轨道周期为基准的起伏状态,因此周期划分参考轨道周期的影响,设置滑动窗口进行分段。综上,根据卫星时序数据的特性,提出一种基于极值的伪周期时间序列分段(EPA_Segment)算法。

2.1.2 算法流程

EPA_Segment 算法包括极值点获取(EPA)和分段(Segment)两部分。该算法根据参数特征选择极大值或者极小值进行处理,处理过程中选择需保持一致,最后通过分析不同选择的分段结果选出最优结果。

极值点获取(EPA)算法基于极值法获得极值点集合,算法流程如表2所示。

分段(Segment)算法基于 EPA 获取的极值点集合EPS进行计算,算法流程如表3所示。

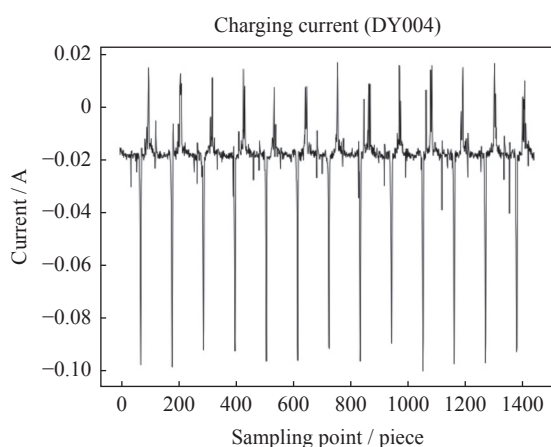


图2 某卫星电源分系统充电电流时序数据

Fig. 2 Charging current timing data of a satellite power supply subsystem

综上,利用 EPA_Segment 算法的两个步骤,通过两次遍历得到时间序列的最优分割点,将原始序列动态地划分为长度不同的周期子序列集合。

2.2 多元参数符号化表示

2.2.1 多元参数异常检测分析

基于多元工程参数进行异常检测,可以归为一个子系统的异常检测。进行单元参数异常检测是多元参数检测的基础,而多元参数异常检测的关键是特征参数的选择和融合。SAX 是一种经典的符号化表示方法,对单元参数时序数据进行符号化表示,有效降低数据维度。但是经典 SAX 算法只能直接对一元时序数据进行符号化表示。另外,SAX 在将数据符号化时,只用均值表示特征,会造成较大的信息损失。 n 元语法模型(N-Gram)是一种基于 $(n-1)$ 阶马尔科夫链的概率语言模型,被广泛用于文本分类和语音识别领域。该模型克服了常规算法中对语言的依赖性,能够提取出潜在特征,不仅可以提高计算机运算效率,还可以利用字符串之间的相似性提取文本特征。

综上,多元参数符号化表示算法(MSAX_NG)设计如下。

(1)使用符号聚合近似算法 SAX 对选中的 m 元参数集进行符号化。

(2)对符号化后的参数集采用 N-Gram 算法,结合上下文信息进行特征融合。

2.2.2 单元参数符号化

经典 SAX 算法可以将任意长度为 n 的时间序列,

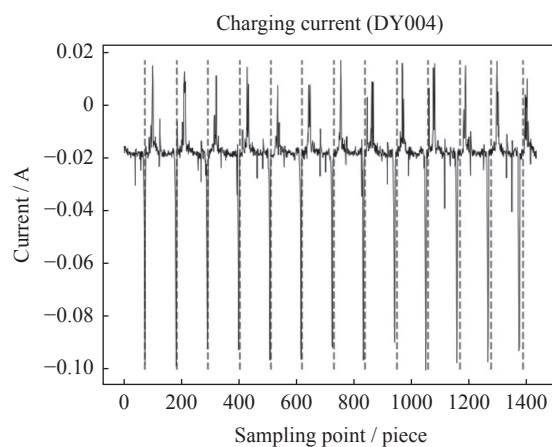


图3 传统的固定窗口子序列分段算法的分段结果

Fig. 3 Segmentation results of the traditional segmentation algorithm for fixed window opening sequence

表 2 极值点获取(EPA)算法流程
Table 2 Extremum point acquisition algorithm process

Input: Time series $T = (t_1, t_2, \dots, t_n)$, threshold
Initialize: The sliding window $\omega = T_{\text{cycle}} + T_{\text{fitted}}$,
 $t_i = t_1$

1. **while** $i \leq n$ **do**
2. $t_i = \text{maximum/minimum}([t_i, t_{i+\omega}])$
3. $t_j = \text{maximum/minimum}([t_{i+\omega-\text{threshold}}, t_{i+\omega+\text{threshold}}])$
4. Add i and j to the extreme point set EPS
5. Set $i = j$
6. **end while**
7. **return** EPS

表 3 分段(Segment)算法流程
Table 3 Segment algorithm process

Input: The extrepoint set sequence EPS = $(e_1, e_2, \dots, e_m)$,
Time series $T = (t_1, t_2, \dots, t_n)$
Initialize: The sliding window size $len, e_i = e_1$

1. **while** $i \leq m$ **do**
2. $t_{e_i} = \text{maximum/minimum}([t_{e_i}, t_{len+e_i}])$
3. $t_{e_j} = \text{maximum/minimum}([t_{e_i}, t_{len+e_i}])$
4. If $e_j = e_i$
5. Add e_i to the set of segmentation time points STP
6. set $e_i = e_j$;
7. **end while**
8. STP = $(s_1, s_2, \dots, s_k)$
9. **while** $i < k$ **do**
10. Add $[t_{s_i}, t_{s_{i+1}}]$ to the set of pseudo-periodic subsequences PPS
11. **end while**
12. **return** PPS

减少到任意长度 w (通常情况下 $w \ll n$)。已知长度为 n 的一元时间序列数据为 $T = (t_1, t_2, \dots, t_j, \dots, t_n)$ 。使用 SAX 进行符号表示的步骤如下。

(1)子序列提取。将长度为 n 的 T 使用滑动窗口 l 和移动步长 p ,提取子序列集合为

$$S_{\text{eg}}(T, l, p) = \{S_1, S_2, \dots, S_{\frac{n-l}{p}+1}\}.$$

(2)子序列标准化。将子序列 S_i 标准化为均值为0、标准差为1的标准序列 S'_i ,即

$$S'_i = (t'_{(i-1) \times p+1}, t'_{(i-1) \times p+2}, \dots, t'_{(i-1) \times p+l}).$$

(3)子序列降维。将标准化后的子序列 S'_i 划分为 k 个截断,取每个截断内参数数据的均值表示该截断,有

$$\bar{S}'_i = (\bar{s}_1, \bar{s}_2, \dots, \bar{s}_k).$$

(4)子序列符号化。选择字母表大小为 φ ,根据字母表大小在表1中查找分裂点 β_i ,再根据图2将均值 $\bar{s}_1$ 表示为对应的字母 $\hat{s}_1$ 。以此类推,将每个子序列符号化为字符串 $\hat{S}'_i$,即

$$\hat{S}'_i = \{\hat{s}_1, \hat{s}_2, \dots, \hat{s}_k\}.$$

(5)最终将时间序列 T 表示为 T' ,有

$$T' = (\hat{S}'_1, \hat{S}'_2, \dots, \hat{S}'_{(n-l)/p+1}).$$

2.2.3 多元参数融合

已知 m 元时间序列长度为 n 的数据序列 $A = (T_1, T_2, \dots, T_i, \dots, T_n)$,其中 T_i 表示第 i 元时间序列数据,算法设计如下。

(1)将 m 元参数按照单元参数符号化表示算法进行符号化表示,分别符号化为 w 个字符串,形成一个 $m \times w$ 阶字符串矩阵,有

$$A' = \begin{bmatrix} \hat{S}'_{11} & \hat{S}'_{12} & \dots & \hat{S}'_{1w} \\ \hat{S}'_{21} & \hat{S}'_{22} & \dots & \hat{S}'_{2w} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{S}'_{m1} & \hat{S}'_{m2} & \dots & \hat{S}'_{mw} \end{bmatrix}.$$

(2)为减少符号化造成的信息损失,避免只考虑单个时间点信息而忽略了上下文信息,每个字符串序列的信息被综合表示为包含上下文信息的矩阵。 u_j 表示符号化后第 j 个 m 元字符串序列,即

$$u_j = [\hat{S}'_{1j}, \hat{S}'_{2j}, \dots, \hat{S}'_{mj}]^T.$$

则可得到第 j 个 m 元字符串序列的上下文信息矩阵,有

$$U_j = [u_{j-1} \quad u_j \quad u_{j+1}] = \begin{bmatrix} \hat{S}'_{1j-1} & \hat{S}'_{1j} & \hat{S}'_{1j+1} \\ \hat{S}'_{2j-1} & \hat{S}'_{2j} & \hat{S}'_{2j+1} \\ \vdots & \vdots & \vdots \\ \hat{S}'_{mj-1} & \hat{S}'_{mj} & \hat{S}'_{mj+1} \end{bmatrix}.$$

(3)采用 1-Gram 表示方法,通过一个大小为1的滑动窗口按顺序从左到右、从上至下对每个 U_j 进行划分,每个 U_j 划分成为一个 gram,统计每个 gram 出现的次数,得到其频率,形成 A_j 的特征向量区间,有

$$U'_j = \{(g_1, y_1), (g_2, y_2), \dots, (g_k, y_k)\}.$$

其中, g_k 表示字符串, y_k 表示字符串出现的频率。

(4) 将 m 元时间序列长度为 n 的数据序列 A 表示为一个一维、频率长度为 w 的字符串序列 B , 即

$$B: \{(B_1, y_1), (B_2, y_2), \dots, (B_w, y_w)\}.$$

3 实验设计

采用的基于 EMF(EPA_Segment, MSAX_NG 和 FastDTW) 的异常检测算法流程如图 4 所示, 其主要步骤包括数据获取、数据预处理、数据伪周期划分、异常检测建模以及实时工程数据异常检测。首先, 获取某卫星的真实工程参数数据, 对数据进行预处理; 进行子序列划分, 生成伪周期序列集合, 对集合进行建模并根据测试集结果调整模型参数, 得到表现最优的模型; 最后, 根据真实卫星的实时数据, 基于模型结果进行异常检测。

3.1 实验数据集

从数据驱动的角度研究多维工程参数进行异常检测, 选择的试验数据均是来自某卫星的真实工程参

数数据。选择的异常数据为值班过程中卫星遇到的真实异常。根据所选择的异常数据选择试验数据, 并按照分系统、参数包、工作模式的形式分别进行建模处理。

3.2 数据预处理

数据预处理过程主要包括异常值剔除和缺失值填补两部分。

3.2.1 异常值剔除

剔除异常值过程主要是针对训练数据进行的。训练数据不包含异常数据, 但是其中可能出现异常值, 必须进行筛选。从预定义的参数监测文件读取参数的阈值上限 U 和下限 L 。对于原始输入数据, 根据阈值上限 U 和下限 L 判断每一时刻点的数据是否需要删除。

3.2.2 缺失值填补

数据缺失常发生在数据采集、传输、存储等过程中。缺失是随机发生的, 目前对于缺失值的处理方法基本分为三种, 即删除、填充、不处理。对于卫星时序数据, 直接删除会造成数据丢失, 不处理则影响后续建模, 因此采用填充的方式处理。

卫星参数具有一定的平稳性, 且所采用的参数数据采样率较高, 短时间内通常不会出现太大波动, 因此直接将前一个时间点的值填补为当前缺失点的值。

3.3 基于 EPA_Segment 的多元卫星工程参数异常检测方法

3.3.1 周期划分

本文算法适用于具有周期特征的参数集, 首先需要分析待检测卫星工程参数集是否具有周期特征, 这主要通过经验进行检验, 绘制参数数据序列示意图(见图 2), 分析参数是否具有周期性特征。对于具有周期特征的数据, 利用 EPA_Segment 算法对参数数据进行周期划分, 生成 N 个伪周期序列。

3.3.2 特征选择

特征选择使用灰色关联度分析法筛选与系统运行具有较高相关性的参数集。灰色关联度分析法需确定反映系统行为特征的主参数序列和影响系统行为的比较参数序列。特征选择流程如图 5 所示。每次建模时, 分析异常出现的特点, 结合先验知识从参数集内选出最能够体现系统特点的参数作为主参数。

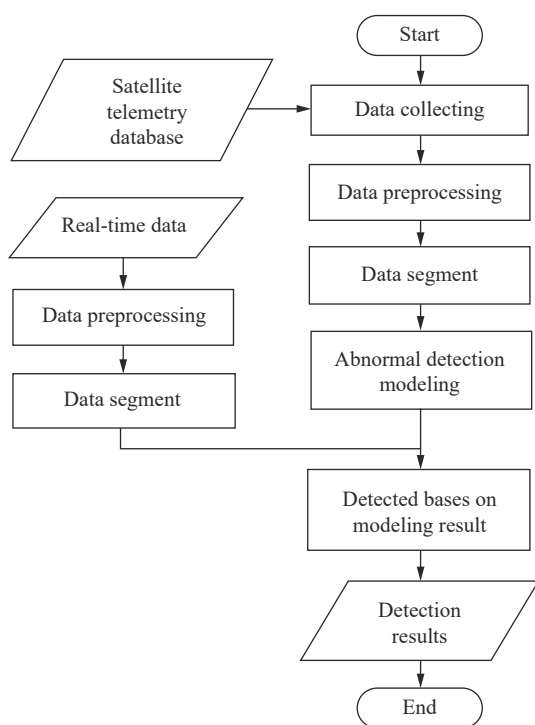


图 4 基于 EMF 卫星异常检测算法流程

Fig. 4 Flow chart of satellite abnormality detection algorithm based on EMF

3.3.3 相似度建模

动态时间规整(DTW)算法^[7]基于动态规划思想,寻找待匹配模板与参考模板之间非线性映射的最优路径。其最大的优点在于可匹配长度不同的序列,适用于具有伪周期特性的卫星工程参数数据。

DTW 算法计算序列所有分量之间的距离,但在路径匹配过程中不会完全用到。Fast-DTW 是 DTW 常用的加速手段,主要通过加约束和数据抽象两种方法加速 DTW 的计算。Fast-DTW 计算相似度不如 DTW 算法准确,但其时间复杂度大大减少,为 $O(n)$ 量级,在进行长时间序列匹配过程中,可大大减少匹配时的计算量和所占用系统空间,有效提高异常检测效率。因此,这里选择 Fast-DTW 算法进行相似性度量。

对多元参数符号化后形成的 N 个字符串集合,利用 Fast-DTW 算法计算各周期序列之间的相似度。每个序列与其他 $N-1$ 个序列计算相似度,形成 $N \times N$ 阶相似度矩阵。计算每一个周期序列与其他周期序列相似度的方差,挑选出方差最小的周期序列作为标准序列,在异常检测过程中与待检测序列进行相似度计算。对于 Fast-DTW 算法,距离越小,相似度越高,因此定义落入 $(\mu+3\sigma, +\infty)$ 区间的序列为异常序列。计算标准序列相似度的均值及方差以确定异常区间。

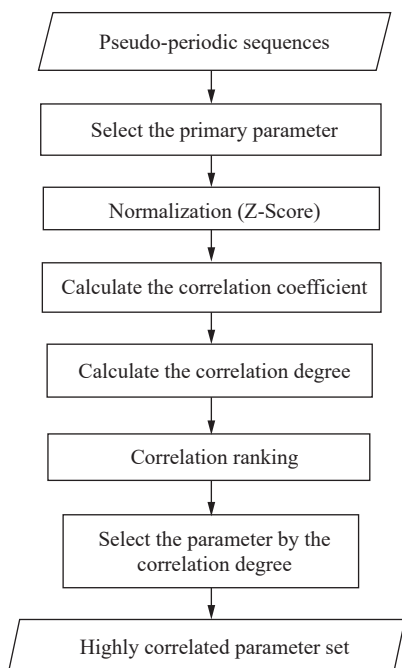


图 5 特征选择流程

Fig. 5 Feature selection flowchart

3.4 检测结果评价

由于数据集的类别不平衡,即正常样本多而异常样本少,异常检测问题的评价指标一般较复杂,不能只计算正确率。在异常检测问题中常用的评价指标包括召回率 Recall 和精确率 Precision,比准确率更加客观。此外,还引入 F1 score 作为综合评价指标。召回率体现了检测的全面程度,精确率体现了检测的准确程度,一定程度上二者越大越好,但是这两个指标通常具有制约效果,在异常检测的评判过程中,保证精确率的情况下提高召回率。F1 指标作为二者的调和均值能够更加综合地表征结果。

4 实验分析

4.1 实验数据分析

实验数据选取某在轨卫星在某月 1 至 31 日两种类型的真实工程参数数据,共得到电源分系统电源主要包相关参数数据 56 维 43253 条,卫星 SAA 数据 4 维 140 条。其中,电源相关数据包含电源状态相关参数(充电电流、母线电压、电源电压等)共 22 维,有效载荷配电参数(星敏感器配电状态、光纤陀螺配电状态等)共 34 维。在数据频率采集方面,电源相关参数为卫星全周期在轨数据,采样间隔为 60 s;在异常数据方面,选取的数据周期内卫星出现若干次真实异常报警情况。根据异常报警情况,将 1 至 11 日的数据定义为训练集,12 至 31 日的数据定义为验证集。经过数据预处理,删除了参数集中仅表示载荷开关状态的参数,最终得到实验数据训练集 48 维 14400 条,验证集 48 维 28853 条。

4.2 周期划分结果与分析

4.2.1 EPA_Segment 参数训练

利用 EPA_Segment 进行周期划分,获得伪周期序列集。分析算法中, EPA_Segment 需要训练参数移动步长 l_{en} 得到最优分段结果。这里选择 $\{2, 3, 4, 5, 6, 7, 8\}$ 进行训练。

4.2.2 EPA_Segment 算法参数训练实验结果

参数训练实验结果列于表 4。由表 4 可知,基于极小值分段计算时,移动步长 l_{en} 为 5 和 6 时分段效果最好。

选择移动步长为 5 时的计算结果,其实验结果如图 6 所示,可以明显看出, EPA_Segment 能够合理地对伪周期序列进行划分。

4.3 特征选择结果与分析

4.3.1 特征选择算法参数训练

采用灰色关联度分析法进行特征选择。根据经验,在电源系统中,选定充电电流作为主参数。算法需要训练的参数为参数选择量(selectparametersnums)。这里选择{3, 4, 5, 6, 7, 8}进行训练。

4.3.2 特征选择算法参数对实验结果的影响与分析

参数选择量对异常检测结果指标的影响如图 7 所示。Recall 指标表现较为敏感,其增加到 7 后不会再影响指标。通过综合分析,以参数选择量 $S=7$ 作为最优特征参数。

电源分系统特征选择结果列于表 5。加上主参数最终得到的数据集共包含 7 维参数。

4.4 异常检测结果与分析

4.4.1 MSAX_NG 参数训练

分析 MSAX_NG 算法,其训练参数包括字母数

表 4 EPA_Segment 算法调参优化结果
Table 4 Results of EPA_Segment tuning optimization algorithm

l_{en}	EPA_Segment (max)/ cycle	EPA_Segment (min) / cycle	Actual/ cycle
2	43	35	12
3	22	20	12
4	18	16	12
5	13	12	12
6	12	12	12
7	12	11	12
8	11	10	12

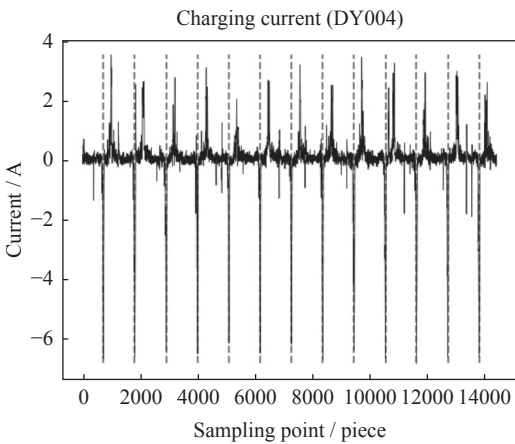


图 6 l_{en} 为 5 时 EPA_Segment 算法的分段结果
Fig. 6 Segment results of EPA_Segment algorithm when l_{en} is 5

φ 、滑动窗口长度 l_{en} 、移动步长 l_{st} 和截断数 k 。

这里选择截断数 $k = \{5, 6, 7\}$; 字母数 $\varphi = \{3, 4, 5\}$; 滑动窗口长度 $l_{en} = \{60, 120, 240, 480\}$; 对于移动步长 l_{st} , 选择为滑动窗口的{0.5, 0.7, 0.9}倍。

训练参数使得找到一组最佳参数值集合 $\{\varphi, l_{en}, l_{st}, k\}$, 得到最好的评价指标结果。

4.4.2 MSAX_NG 算法各项参数的影响

按照 MSAX_NG 参数训练过程,在训练集上进行实验,各评价指标结果如图 8 所示。从图 8 (a) 可以看到,截断数 k 从 5 增长到 7,各个指标呈现先增加后减小的趋势,在 $k=6$ 时各个指标表现最好;从图 8 (b) 可以看到,滑动窗口长度 l_{en} 从 60 依次加倍至 480,各个指标呈现短暂增加后快速降低,尤其是 recall 在 120 之后下降比较明显,而 precision 则有缓慢提升后再下降,这主要是因为滑动窗口过小会分割异常数据段,造成其异常特征表现减弱,过大会使得异常数据与正常数据段混合而造成召回率降低,在 $l_{en}=120$ 时各个指标综合表现最好;从图 8 (c) 可以看到,移动步长从 30 增加至 240,会明显影响召回率,而对精确度的影响不明显,在步长为 72 时各个指标表现最好。对于字母数 φ ,从 4 到 5,通常最好情况下各个指标在 $\varphi=5$ 时比 $\varphi=4$ 时表现更好,但 $\varphi=5$ 时更容易受到滑

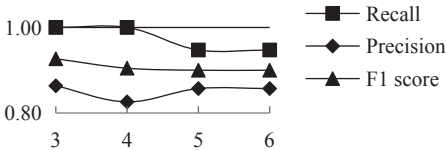


图 7 参数选择量对异常检测评价指标的影响
Fig. 7 Effect of selectparametersnums on the evaluation index of anomaly detection

表 5 电源分系统特征选择的最优结果
Table 5 Optimal result of the feature selection of the power supply subsystem

卫星参数	相关性百分比/(%)
测控应答机A机配电状态	83.5
测控应答机B机配电状态	83.3
GNSS接收机A机配电状态	83.0
备相位计配电状态	82.9
备载荷电控箱配电状态	82.8
相位计配电状态	82.8

动窗口和移动步长的影响,因此稳定性不如 $\varphi=4$ 时表现好。

综上,选择 $\{\varphi=4, l_{\text{en}}=120, l_{\text{st}}=72, k=6\}$ 作为最佳参数集进行模型验证。

4.5 对比算法与实验结果

选择有监督的学习算法 1-NN ED 和无监督学习算法 KMeans ED 进行比较,其结果列于表 6。

对各个算法的 3 个评价指标进行统计,结果显示本文使用的 EMF 算法在各个指标上均优于 1-NN ED 和 KMeans ED 算法,其表现最佳。综合评价,算法表现效果的排序为 $\text{EMF} > 1\text{-NN ED} > \text{KMeans ED}$ 。

KMeans ED 表现不如 1-NN ED,但是其属于无监督学习算法,适用于检测未知的异常。1-NN ED 与 EMF 算法精确率近似,但是 EMF 的召回率明显优于 1-NN ED,表示在异常检测过程中具有更全面的异常检测能力。

5 结论

本文提出了 EMF 算法,首先分析卫星工程参数数据特性,对具有周期特征的参数集,利用 EPA_Segment 算法将数据划分为伪周期序列,再利用 MSAX_NG 算法得到符号化表示及特征融合后生成的字符串集合,基于此使用 Fast-DTW 算法进行相似性度量并建立异常检测模型,最后利用在轨运行的某型号科学卫星工程参数数据进行模型验证。

实验结果表明,EMF 算法能够将参数序列有效划分为伪周期序列,并且在实际卫星异常检测验证结果中,其在 recall, precision 和 F1 score 评价指标上均具有较好的表现。此外,分析分类算法检测出的异常,发现本算法对于组合异常表现优异。

实验过程优化训练算法的参数并分析其对异常检测结果的影响:特征选择过程中参数选择数量并非越多越好,在某个值之后其对检测结果的影响很小并且会增加模型的复杂程度;MSAX_NG 中的参数字母数 ϕ 、滑动窗口长度 l_{en} 、移动步长 l_{st} 和截断数 k 均会明显影响检测结果,需对不同的参数进行调优获得最好的模型。

本文提出的 EMF 算法利用在轨运行的某型号科学卫星工程参数数据进行建模以及模型验证,结果表明其能够应用于实际卫星故障检测,呈现良好的检测

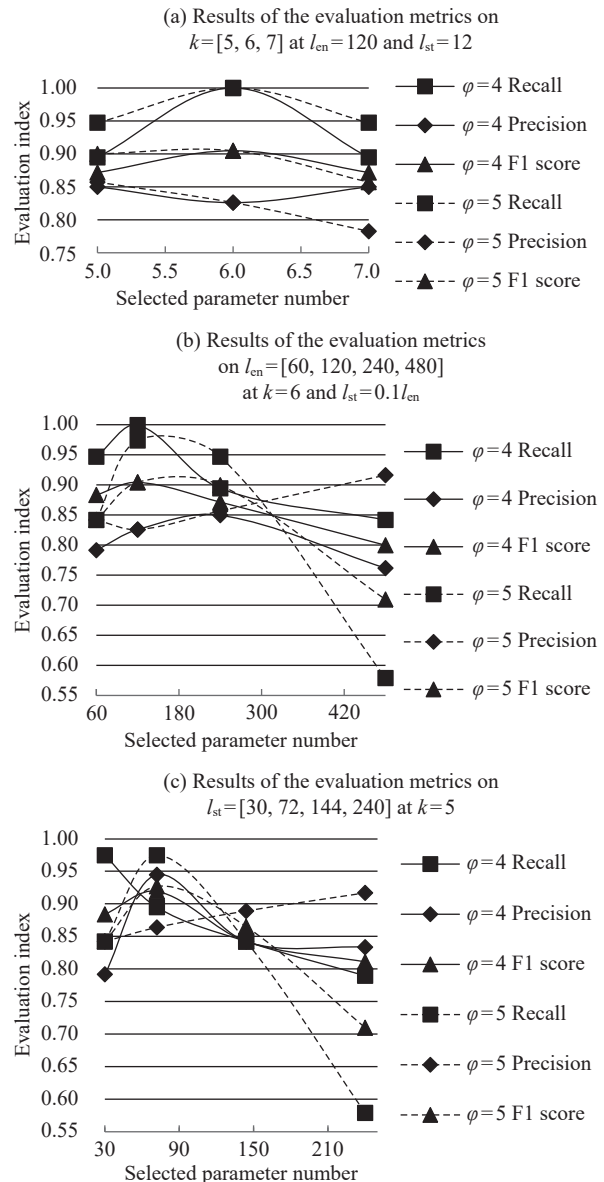


图 8 参数训练过程中各评价指标结果

Fig. 8 Results of each evaluation index during the parametric training process

表 6 不同算法的异常检测结果
Table 6 Anomaly detection results for different algorithms

评价指标	EMF	1-NN ED	KMeans ED
Recall	0.947	0.871	0.833
Precision	0.9	0.9	0.714
F1 score	0.923	0.885	0.769

结果,具有工程应用意义。在未来的研究中,有以下问题需做进一步探讨和改进:(1)在针对卫星伪周期

序列特性的研究中,仅考虑了卫星轨道对周期特性的影响,后续可增加太阳、空间任务的影响;(2) EMF 算法还可与其他异常检测算法(例如聚类算法、神经网络等)相结合,并探索其对异常检测效率的影响;

参考文献

- [1] ZHANG Xiaofang, LIU Songtao, WU Yaoping. Statistical analysis of space weather effects on satellites anomalies[J]. *Chinese Journal of Space Science*, 2015, **35**(4): 461-472 (张晓芳, 刘松涛, 吴耀平. 影响卫星故障的空间天气分析[J]. 空间科学学报, 2015, **35**(4): 461-472)
- [2] PANG Jingyue. Probabilistic Prediction Based Anomaly Detection Method for Spacecraft Telemetry Data[D]. Harbin: Harbin Institute of Technology, 2019 (庞景月. 基于概率性预测的航天器遥测数据异常检测方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2019)
- [3] YANG Jiasen, MENG Xin, CHEN Tuo, *et al.* Anomaly detection of spacecraft based on the telemetry data correlation[J]. *Chinese Journal of Scientific Instrument*, 2018, **39**(8): 24-33 (杨甲森, 孟新, 陈托, 等. 基于遥测数据相关性的航天器异常检测[J]. 仪器仪表学报, 2018, **39**(8): 24-33)
- [4] HUNDMAN K, CONSTANTINOU V, LAPORTE C, *et al.* Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding[C]//Proceedings of the 24 th ACM Sigkdd International Conference on Knowledge Discovery & Data Mining. London: ACM, 2018: 387-395
- [5] ZHANG Wei, XIAO Zhigang, LI Boquan. Application of fault tree expert system in diagnosis about payload status[J]. *Computer Technology and Development*, 2014, **24**(12): 28-31,36 (张薇, 肖志刚, 李博权. 故障树专家系统在有效载荷状态诊断的应用[J]. 计算机技术与发展, 2014, **24**(12): 28-31,36)
- [6] FIGUEROA F, UNDERWOOD L, WALKER M G, *et al.* NASA platform for autonomous systems (NPAS)[C]//AIAA SciTech. San Diego: AIAA, 2019
- [7] BERNDT D J, CLIFFORD J. Using dynamic time warping to find patterns in time series[C]//Proceedings of the 3 rd International Conference on Knowledge Discovery and Data Mining. Seattle: ACM, 1994: 359-370
- [8] TIAN Linlin. Spacecraft fault detection technology based on deep learning and GPU computing[J]. *Computer Measurement & Control*, 2020, **28**(5): 1-4,9 (田林琳. 基于深度学习及GPU计算的航天器故障检测技术[J]. 计算机测量与控制, 2020, **28**(5): 1-4,9)
- [9] GU Xinyu. Research on Satellite Telemetry Data Prediction based on ARIMA-SVR Combination Model[D]. Beijing: University of Chinese Academy of Sciences (National Space Science Center, Chinese Academy of Sciences), 2021 (顾昕雨. 基于ARIMA-SVR组合模型的卫星遥测数据预测研究[D]. 北京: 中国科学院大学(中国科学院国家空间科学中心), 2021)
- [10] WANG Wei, WANG Jiandong, ZHANG Xia. An anomaly detection method of airport-noise time series based on improved SAX measurement[J]. *Computer and Modernization*, 2014(8): 5-10 (王伟, 王建东, 张霞. 基于改进符号化度量方法的机场噪声异常检测[J]. 计算机与现代化, 2014(8): 5-10)
- [11] RUAN Hui, LIU Lei, HU Xiaoguang. Satellite time series data classification method based on trend symbolic aggregation approximation[J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2021, **47**(2): 333-341 (阮辉, 刘雷, 胡晓光. 基于趋势符号聚合近似的卫星时序数据分类方法[J]. 北京航空航天大学学报, 2021, **47**(2): 333-341)
- [12] WANG Jian, MAO Liming, YIN Aijun. Multi-dimensional DTW combined with shape feature and context information[J]. *Computer Engineering and Applications*, 2020, **56**(22): 42-47 (王见, 毛黎明, 尹爱军. 结合形状特征及其上下文的多维DTW[J]. 计算机工程与应用, 2020, **56**(22): 42-47)
- [13] DU Ying, LI Sihui, HU Shaolin, *et al.* Detection algorithm of abnormal sub-sequence in pseudo periodic time series of telemetry[J]. *Chinese Space Science and Technology*, 2016, **36**(5): 18-24 (杜莹, 李四虎, 胡绍林, 等. 遥测伪周期时间序列子序列异常检测算法[J]. 中国空间科学技术, 2016, **36**(5): 18-24)
- [14] LIN J, KEOGH E, WEI L, *et al.* Experiencing SAX: a novel symbolic representation of time series[J]. *Data Mining and Knowledge Discovery*, 2007, **15**(2): 107-144